The Impact of Market Growth on Delivery Time: Evidence from JD.com

Journal:	*Production and Operations Management*
Manuscript ID	POM-Mar-23-OA-0233
Wiley - Manuscript type:	Original Article
Keywords:	market growth, delivery speed, e-commerce, quality of service, geographical expansion

Abstract:	Market growth can offer significant advantages for companies in terms of economies of scale and operational efficiencies, especially when it is accompanied by an increase in customer density. However, market expansion that is driven by geographically dispersed demand can create negative trade-offs for e-commerce platforms. In this study, we examine the impact of market growth on delivery speed using data provided by JD, a prominent e-commerce platform in China. Our analysis indicates that newer JD customers experience progressively slower deliveries than legacy customers, with each additional year on the platform resulting in a 0.67 hour decrease in average delivery time. These findings suggest that companies may have an additional lever to improve customer acquisition and retention, beyond just offering financial incentives. Specifically, to enhance the retention of newer customers who are not yet loyal to the business, online platforms may want to prioritize faster processing and delivery for these customers.

The Impact of Market Growth on Delivery Time: Evidence from JD.com

(Authors' names blinded for peer review)

Market growth can offer significant advantages for companies in terms of economies of scale and operational efficiencies, especially when it is accompanied by an increase in customer density. However, market expansion that is driven by geographically dispersed demand can create negative trade-offs for e-commerce platforms. In this study, we examine the impact of market growth on delivery speed using data provided by JD, a prominent e-commerce platform in China. Our analysis indicates that newer JD customers experience progressively slower deliveries than legacy customers, with each additional year on the platform resulting in a 0.67 hour decrease in average delivery time. These findings suggest that companies may have an additional lever to improve customer acquisition and retention, beyond just offering financial incentives. Specifically, to enhance the retention of newer customers who are not yet loyal to the business, online platforms may want to prioritize faster processing and delivery for these customers.

Key words: Supply chain, market growth, delivery speed, e-commerce, quality of service, geographical expansion

1. Introduction

Online e-commerce platforms such as Amazon, Alibaba, and JD have enjoyed considerable market growth in recent years. Their growth has accelerated during the COVID-19 pandemic, with gross merchandise value up 81% year over year in the fourth quarter of 2020 (WSJ, 2021). Some expect that these dynamics can be sustained as we usher in a new era of e-commerce dominance post-pandemic (Deloitte, 2021).

Market growth is a key driver of success for many businesses, as it enables them to achieve economies of scale and operational efficiencies (Deloitte, 2015). However, the impact of market growth on e-commerce platforms can be complex, particularly when it is driven by geographically dispersed demand. In such cases, the trade-offs between market growth and efficient delivery can be challenging to manage. Could there also be some negative trade-offs with market growth in e-commerce supply chains that would counteract the well-known benefits of economies of scale? If so, how do they affect customers?

To gain insights into the impact of market growth on delivery speed, we investigate these questions using transaction and logistics data provided by JD[1] (Shen et al., 2020). We find that customers who have joined the platform more recently (newer customers) suffer from progressively slower deliveries relative to legacy customers who have been using the platform longer. For example, we find that for each additional year a customer has been on the platform, the average shipping time decreases by around 0.61 hours and the probability of the package not arriving within 24 hours decreases by about 1.9%.

These findings might seem to contradict the literature on dynamic competitive environments that suggests that firms may want to first provide a bargain (e.g., a lower price) to attract and "lock-in" new customers (e.g., Klemperer, 1995; Padilla, 1992; Chen, 1997; Anderson and Simester, 2004; Farrell and Klemperer, 2007). Then, once the customers are "locked-in", the seller can take advantage of them to extract larger surplus. A similar approach with regard to delivery speed may seem to be a sensible strategy for an e-commerce platform that would benefit from offering newer customers faster deliveries, possibly at the expense of locked-in legacy customers. However, our results suggest that other mechanisms restrict firms' ability to deliver on such a strategy despite the economies of scale.

In particular, our results suggest that JD's market growth is currently being driven by harder-to-serve customers joining the platform from geographically distant areas with potentially less-developed distribution infrastructure. This hypothesis aligns with the general growth pattern of Chinese e-commerce firms: start in densely populated areas, and after demand there is tapped, gradually extend into rural areas (Wong and Chao, 2015; Watanabe, 2019; Shi et al., 2017; and Lee, 2009) that become the new center of demand growth, as shown in Figure 1.

We believe that this expansion into rural areas is what causes newer customers to receive slower shipping. To support this claim, we provide empirical evidence that newer customers are harder to reach. First, we show that newer customers are less likely to receive packages from the warehouse closest to them. Specifically, we find that each additional year a customer has been on the platform increases the probability that the nearest warehouse fulfills the order by about 1.45%. This finding suggests that newer customers tend to reside near

[1] JD is a leading e-commerce platform in China with net revenue in 2020 of $114.3bn and double-digit growth in its customer base (JD, 2021).

smaller, less developed warehouses. Second, we show that newer customers receive slower shipments even when they are served by the closest warehouse. Specifically, we find that each additional year a customer has been on the platform decreases the shipping time for shipments made from the nearest warehouse by around 0.18 hours. This finding suggests that newer customers tend to be harder to serve due to less-developed local supply chain infrastructure.

The logistical trade-off between market growth and efficient delivery is not unique to JD in China. E-commerce platforms around the world face similar challenges as they expand their operations and customer base. The problem arises when a company's customer base becomes geographically dispersed, and the delivery times increase as a result of longer distances and increased shipping costs. This can negatively impact the customer experience, leading to dissatisfaction and churn.

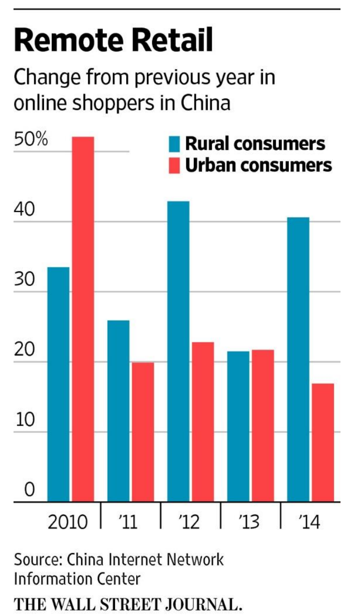

Figure 1 **E-commerce growth in rural areas has eclipsed main cities (Wong and Chao, 2015).**

However, it is not always feasible for e-commerce platforms to establish a physical presence in all demand regions immediately. Doing so requires significant investments in logistics infrastructure and operations, which may not be financially feasible or logistically possible for a company. Therefore, companies must balance the benefits of market growth with the challenges of ensuring efficient delivery.

Our findings highlight the importance of balancing market growth with efficient delivery, and the need for e-commerce companies to adopt innovative strategies to manage the trade-offs between these competing objectives. By prioritizing packages to newer, hard-to-reach customers, online platforms can achieve sustainable growth that mitigate the negative effects of geographically dispersed demand. Instead of aggressively targeting new customers with financial incentives and promotions, companies can prioritize faster processing and delivery times for new customers to lock-in their loyalty and retention. This can be achieved by leveraging advanced data analytics and machine learning algorithms to identify the most profitable customers and optimize the allocation of logistics resources.

2. Literature review

Demand growth is often associated with benefits arising from economies of scale: firms can reduce their marginal operating and capacity costs, reduce demand/supply uncertainty, and provide higher-quality service (see, for example, Manne, 1961). In operations, this is often referred to as the benefits of demand pooling (Eppen, 1979). Several empirical studies support these conclusions across various industries. For example, in retail, Rumyantsev and Netessine (2007) analyze data from hundreds of public firms and find evidence suggesting that larger firms carry relatively less inventory. Gaur and Kesavan (2015) use financial data from US retailers to show that firms with higher sales also experience faster inventory turnover.

Similar results can be found in the context of service industries. Classical modeling literature, such as Smith and Whitt (1981) and Whitt (1992), shows that pooling customer queues (e.g., in call centers) might reduce wait time and increase server utilization. In the hospital setting, a greater number of treated patients might decrease per-patient costs and improve the overall quality of treatment (Giancotti et al., 2017).

Nevertheless, some studies highlight potential trade-offs with the increase in scale. Economies of scale are often associated with diminishing returns - the marginal benefits from demand pooling decrease with demand growth (e.g., Anupindi et al., 2012). Gaur and Kesavan (2015) empirically illustrate this diminishing return to scale in retail with regard to inventory turnover: *"inventory turnover increases with size at a slower rate for large firms than for small firms."* At some point, firms may even experience diseconomies of scale (Williamson, 1975). Broadly speaking, once the firm exceeds a certain size threshold, the marginal cost of a product or service begins to increase due to several factors, such as the increasing complexity of the business, lack of communication, etc. For example, in queueing, demand pooling sometimes leads to negative mechanisms that outweigh the benefits of economies of scale, with dedicated queues achieving shorter wait time and higher quality of service compared to pooled queues (e.g., Mandelbaum and Reiman, 1998). Pang and Whitt (2010) argue that large-scale service systems, while benefiting from economies of scale, are more vulnerable to negative effects from disruptions (e.g., increased congestion). In another example, Song et al. (2015) provide empirical evidence that there can be diseconomies of scale in pooling queues at hospitals because physicians tend to exhibit less "ownership" over patients in pooled queues than over patients in dedicated ones.

To the best of our knowledge, our paper is the first to empirically study the impact of market growth on supply chain performance in terms of the quality of service to the end customer. An important quality-of-service metric for e-commerce is delivery speed, which is considered a key competitive advantage (e.g., Cui et al., 2019 and Fisher et al., 2019). The papers investigating quality of service in e-commerce include Cui et al. (2021), who look at how over- and underpromising delivery time affects sales, and Bray (2020), who studies how customers react to shipping notifications when delays occur at different delivery stages. Results in these papers suggest that quality of service in e-commerce can have a noticeable effect on customers.

One way an e-commerce firm can provide faster deliveries is by increasing the density of its distribution network - i.e., by building more warehouses near customers. Doing so can lead to economies of density. In contrast to economies of scale, economies of density arise when, for example, supply chains expand in already existing markets, thereby benefiting from shorter shipping distances and better inventory coordination (e.g., Holmes, 2011 and Houde et al., 2021a). However, there is a trade-off between economies of scale and density, as firms often face a choice between building more warehouses or stores (increasing density) or consolidating them into bigger ones (increasing scale).

In the operations literature, the trade-off between economies of density and scale is studied mainly in the context of facility location problems and their extensions (e.g., Daskin, 2013 and Lim et al., 2017). In the economics literature, Holmes (2011) uses data from Walmart to optimize the location of its stores by taking into consideration the benefits of inventory coordination and shorter shipping distances as well as property and labor costs and demand cannibalization (which occurs when stores are built too close to one another). The author explains that Walmart's strategy of building new stores near existing ones allows the firm to reap benefits from economies of density.

Houde et al. (2021a) study data from Amazon and show that it can realize significant cost savings by expanding the density of its distribution network and reducing the distance between distribution centers and customers. In contrast to our paper, the focus of the above-mentioned work is on cost optimization, not the impact of demand growth on the quality of service to the end customer. In fact, Houde et al. (2021b) mention that Amazon is already able to deliver fast enough to most of its customers, and the proximity of its distribution network does not translate to even faster shipping. We complement this literature by empirically illustrating in the context of JD how demand growth can lead to a heterogeneous delivery speed between newly acquired and legacy customers, thereby creating a negative trade-off with the traditionally known benefits of market growth.

3. Data and variables

The original JD dataset consists of seven tables that describe all customer-level transactions and shipments for a single product category in March 2018 (see Shen et al. (2020) for a detailed description). We use only four of these tables. The first table we use describes the SKUs, of which there are two types: JD and non-JD. The former are owned, managed, and delivered by JD; the latter are managed and, in most cases, delivered by third-party sellers. This table also includes unique SKU and brand identifiers.

The second table we use contains information about customers, including when they first purchased on the platform, how frequently they purchased in the past, and whether they have a premium "plus" subscription.

The third and fourth tables we use contain information about orders and deliveries, including when orders were placed, when shipments left warehouses for last-mile facilities, when they left last-mile facilities for final delivery, and when packages arrived at their final destinations. These tables also list the identification numbers of originating warehouses and final customer residences, as well as other order details, such as item quantities, prices, and discounts.

We merge the above-mentioned data tables to create a sample comprising 486,928 distinct orders. We then exclude the orders

- with at least one non-JD SKU ($\sim 49.3\%$ of the initial sample), because delivery information for most non-JD items is missing;
- with multiple distinct JD SKUs ($\sim 9.8\%$ of the remaining sample), so we can associate each shipment with a single SKU dummy variable;
- made by business customers ($\sim 2\%$ of the remaining sample), since they could order large quantities and/or have special delivery arrangements with JD;
- with missing information about promised delivery time ($\sim 0.6\%$ of the remaining sample), because we use promised delivery time as an outcome variable; or
- with missing actual delivery information ($\sim 1\%$ of the remaining sample).

Our final sample comprises 214,726 orders and the same number of shipments, since each remaining order corresponds to only one shipped package.[2]

3.1. Independent variable and controls

Our unit of analysis is the customer order. Our primary independent variable is the *customer lifetime*, which measures the number of months since a given customer made their first purchase.

And our control variables are as follows:

- *is_plus* is a dummy variable that indicates customers with the premium "plus" membership. Shen et al. (2020) explain that "JD PLUS membership costs up to US $45 per year and members enjoy a variety of perquisites including exclusive discounts, higher purchasing reward rate, free delivery, and return with no pre-conditions."
- *SKU id* is a unique SKU identifier. It has 321 distinct values.
- *quantity* is the total number of items (of the same SKU) in the shipment.
- *order date* is the integer between 1 and 31 indicating the day in March 2018 when the order was placed.
- *order hour* is the integer between 1 and 24 indicating the hour the order was placed.
- *purchase intensity* is a categorical variable created by JD that ranks customers from "1" (lowest) to "4" (highest) based on the total value of their past purchases. In addition, customers who just made their first purchase are assigned value "0". *purchase intensity* for some customers is unknown.
- *origin warehouse id* is a unique identifier of the warehouse from which an order was fulfilled. This variable has 55 unique values.
- *local warehouse id* is a unique identifier of the local warehouse closest to the customer's designated shipping address. This variable has 60 unique values.
- *active local warehouse* is a dummy variable that indicates whether a local warehouse ships any orders in our sample (not all warehouses carry the given SKU category).
- *city size* is a categorical variable created by JD that rates the size of the customer city on a scale of 1 (smallest) to 5 (largest). *city size* for some shipments is unknown.
- *is_gift* is a dummy variable indicating whether a given SKU is a free gift from JD.

Shen et al. (2020) provide a more detailed description of these control variables and their summary statistics.

[2] Given that our data describes only a single product type, it is possible that a shipment includes other items not included in our dataset. In such cases, some shipments can still consist of multiple SKUs.

3.2. Dependent variables

We gauge supply chain speed with several metrics (see Table 1 for summary statistics):

- *order processing time* is the number of hours between when the order is placed and when it leaves the initial warehouse.
- *shipping time* is the number of hours between when the package leaves the initial warehouse and when it arrives at the final destination.
- *delivery time* is the sum of the *shipping time* and *order processing time*.
- *promised delivery time* is the integer between 1 and 8 indicating the number of days pledged to the customer at checkout.
- *multi-day delivery* is a dummy variable that indicates whether the *delivery time* exceeds 24 hours.
- *local shipping* is a dummy variable that indicates whether the *origin warehouse id* matches the *local warehouse id*; i.e., whether the shipment is made from the warehouse closest to the customer.
- *local warehouse assortment* is the average number of unique JD SKUs held at the local warehouse. This variable is equal to zero for the warehouses that did not ship orders.
- *local warehouse volume* is the total number of orders fulfilled by the local warehouse. This variable is equal to zero for the warehouses that did not ship orders.

Table 1 Summary statistics for the outcome variables

	Mean	Standard deviation
customer lifetime (months)	36.55	26.87
order processing time (hours)	5.52	12.06
shipping time (hours)	23.73	20.26
delivery time (hours)	29.25	24.46
promised delivery time (days)	1.65	0.96
multi-day delivery	0.42	0.49
local shipping	0.65	0.48

4. Exploratory data analysis

We will first present some informal exhibits that illustrate that JD ships more slowly to newer customers.

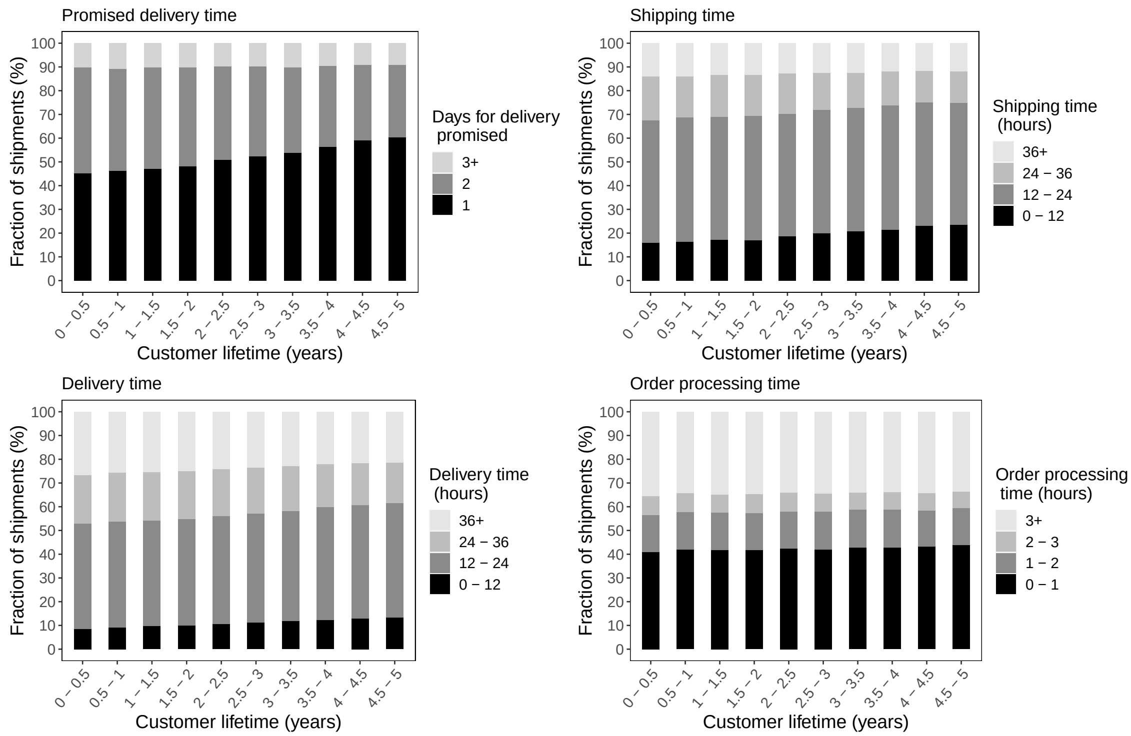

Figure 2 These plots illustrate how the distributions of our dependent variables - promised delivery time, delivery time, shipping time, and order processing time, - vary with customer lifetime. For example, the promised delivery time was one day for about 45.11% of orders placed by customers with a lifetime of less than six months and about 56.37% of orders placed by customers with a 3.5-4 year lifetime.

Figure 2 illustrates how the distributions of shipping, delivery, promised delivery, and order processing times change with the customer lifetime. We observe that customers who are newer to JD are progressively more likely to receive slower deliveries. For example, only 52.81% of orders placed by customers who joined JD's platform within the last 6 months are delivered within 24 hours, in contrast with 60.72% of orders placed by customers who first purchased 4-4.5 years ago. However, it is worth noting that for all customer lifetime values, it takes approximately the same amount of time to dispatch the order from the warehouse once it is received. This suggests that the slower deliveries are due to differences in shipping logistics rather than differences in warehouse operations. For example, the fraction of orders with a shipping time shorter than 12 hours is 15.88% for customers who joined within the last 6 months and 21.47% for customers who joined 3.5-4 years ago.

Figure 3 illustrates that the average shipment is slower for newer customers at 45 of 55 warehouses and across 47 of the 55 best-selling SKUs. Note that if new and legacy customers had the same shipping time distribution, these statistics would be drawn from a

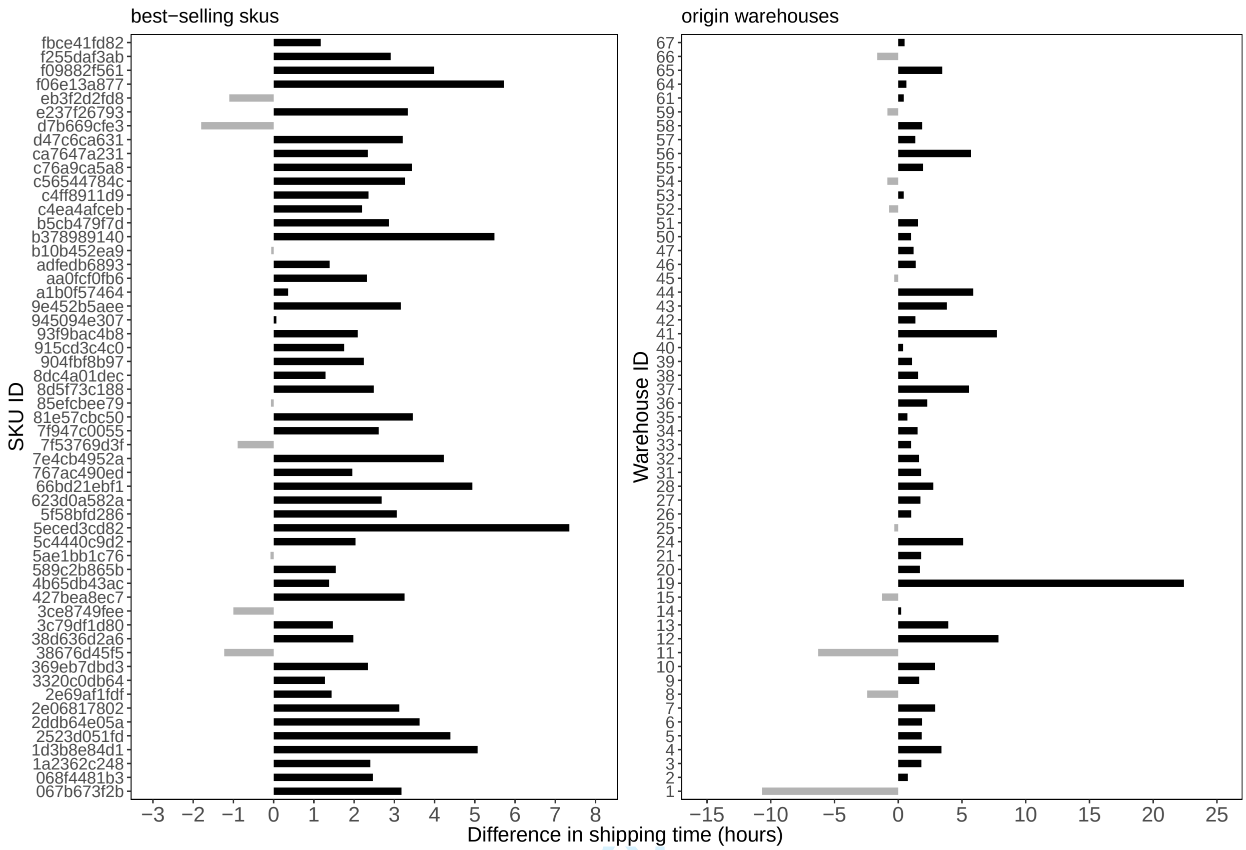

Figure 3 These graphs depict the average shipping time of customers who joined within the last year minus the average shipping time of customers who joined between 3 and 4 years ago. We plot these differences for each of the 55 best-selling SKUs and each of the 55 origin warehouses. Positive black bars indicate that the average shipping time for new customers is longer than the average shipping time for legacy customers. Negative gray bars indicate the opposite. For example, it takes on average more than three hours longer for SKU "067b673f2b" to arrive to a customer who first purchased 0-1 years ago than to arrive to a customer who first purchased 3-4 years ago.

Binomial $(55, 0.5)$ distribution. And there is only a $8.12 \cdot 10^{-7}$ probability that a Binomial $(55, 0.5)$ distribution is as large as 45, and only a $3.38 \cdot 10^{-8}$ probability that it is as large as 47.

Finally, Table 2 demonstrates that newer customers are less likely to receive packages from the local warehouse. For example, customers who first purchased less than 6 months ago are about 2.36% less likely to receive a local shipment than are customers who joined 4-4.5 years ago.

Table 2 Fraction of shipments fulfilled by the warehouse that is closest to the customer by six-month customer lifetime buckets.

customer lifetime (years)	proportion of local shipment (%)
0 - 0.5	63.71
0.5 - 1	63.16
1 - 1.5	63.72
1.5 - 2	63.82
2 - 2.5	64.01
2.5 - 3	64.00
3 - 3.5	64.18
3.5 - 4	64.85
4 - 4.5	66.07
4.5 - 5	65.92
5 - 5.5	68.28
5.5 - 6	67.97

5. Base model

We now use regressions to establish the statistical significance of our results. We first run a set of OLS regressions with the following specification:

$$outcome\ variable\ = customer\ lifetime\ +\ controls, \tag{1}$$

where the outcome variable is either the *shipping time*, *delivery time*, *promised delivery time*, or *multi-day delivery*, and the controls are either nothing or *is_plus*, *is_gift*, *quantity*, *order date*, and *order hour*.

We factor *is_plus* because JD may prioritize faster shipping to its premium customers. We factor *quantity* because larger orders may delay shipments. We factor *is_gift* because JD might not prioritize shipping of free items. We also factor *order date* and *order hour* because newer customers may be more likely to place orders, for example, during weekends or at night.

As a robustness check, we also run a logistic regression of *multi-day delivery* on *customer lifetime* and an ordinal logistic regression of *promised delivery time* on *customer lifetime* with the same two specifications.

Table 3 reports the regression coefficient estimates [3]. The *shipping time*, *delivery time*, *promised delivery time*, and *multi-day delivery* are all significantly negatively correlated with the *customer lifetime* at the $p = 0.01$ level. For example, a one-year decrease in customer lifetime reduces the probability of receiving the package within 24 hours by around 1.9%.

Table 3 This table reports the customer lifetime estimates from the specification (1) regressions. The numbers in parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error for the OLS regressions and the standard error for the logistic regressions.

regression type	outcome variable	without controls	with controls
OLS	shipping time (hours)	-0.0512^{***} (0.00157)	-0.0507^{***} (0.00166)
OLS	delivery time (hours)	-0.0559^{***} (0.002)	-0.0609^{***} (0.00213)
OLS	promised delivery time (days)	-0.00362^{***} $(7.498 \cdot 10^{-5})$	-0.00374^{***} $(8.008 \cdot 10^{-5})$
ordinal logistic	promised delivery time	-0.0108^{***} (0.000163)	-0.0111^{***} (0.000175)
OLS	multi-day delivery (%)	-0.158^{***} (0.00408)	-0.18^{***} (0.00425)
logistic	multi-day delivery	-0.00659^{***} (0.000166)	-0.00815^{***} (0.000185)
Note:		*p<0.1; **p<0.05; ***p<0.01	

To show the ubiquity of this result, we slice the data in seven different ways. Specifically, we run separate OLS regressions of the *shipping time* on *customer lifetime* without controls for

- all 55 warehouses,
- all 258 SKUs that recorded at least 20 orders,
- all 4 levels of the *promised delivery time*,
- all 4 levels of the *purchase intensity*,
- all 5 levels of the *city size*,
- both levels of *plus status*, and
- both levels of *local shipping*.

[3] For all OLS regressions in this paper, we report the maximum of the bootstrapped standard error (that is robust to general heteroskedasticity (Cameron and Trivedi, 2005)) and the HC_1 heteroskedasticity robust standard error (MacKinnon and White, 1985).

Table 4 This table reports the customer lifetime estimates from specification (1) regression for the *shipping time* as the outcome variable run across the 17 subsamples listed in the first column. The numbers in parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error.

promised delivery time	estimate	observations
1 and 2 days	-0.0382^{***} (0.000957)	193, 767
1 day	-0.0102^{***} (0.00119)	115, 774
2 days	-0.0141^{***} (0.00167)	77, 993
3 days or longer	-0.0792^{***} (0.00887)	20, 959
purchase intensity	**estimate**	**observations**
1	-0.0221^{***} (0.00835)	43, 758
2	-0.0334^{***} (0.00453)	62, 914
3	-0.0435^{***} (0.00387)	51, 777
4	-0.0449^{***} (0.00328)	54, 595
city size	**estimate**	**observations**
1	-0.0152^{***} (0.00216)	52, 889
2	-0.0196^{***} (0.00255)	69, 422
3	-0.0148^{***} (0.0044)	31, 342
4	-0.00941^{*} (0.00486)	26, 537
5	-0.0154 (0.0355)	2500
plus customers	**estimate**	**observations**
Yes	-0.0451^{***} (0.00313)	53, 398
No	-0.0557^{***} (0.00195)	161, 328
local shipping	**estimate**	**observations**
Local	-0.0464^{***} (0.00152)	140, 285
Non-local	-0.02^{***} (0.00343)	74, 441
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

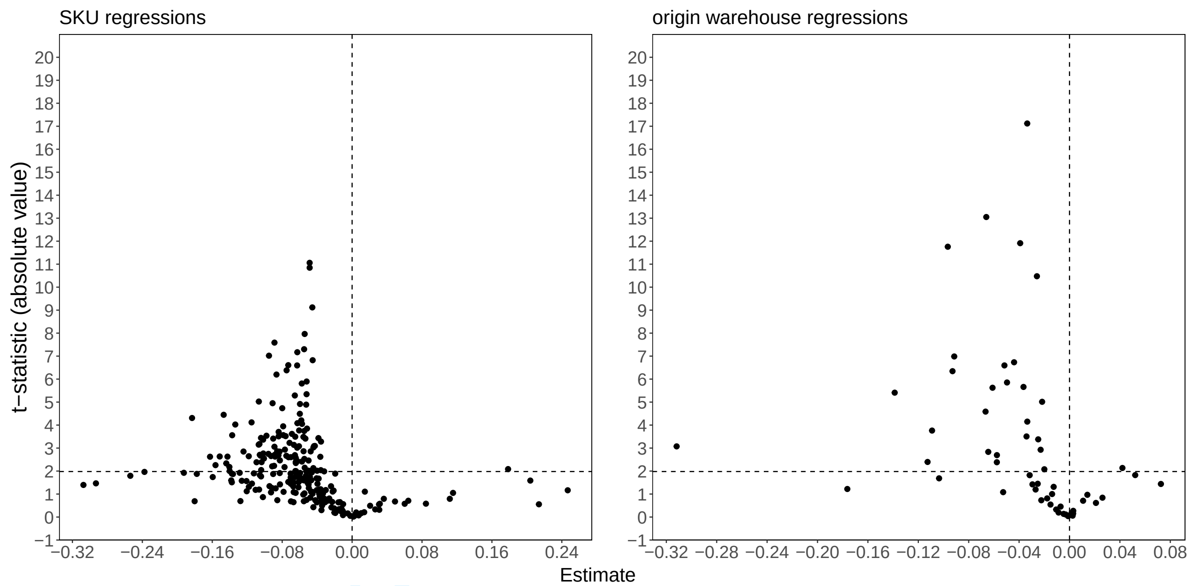

Figure 4 This figure depicts the customer lifetime coefficient estimates and their corresponding t-statistic absolute values for the specification (1) regressions without controls. We run these regressions separately for each of the 55 origin warehouses and each of the 258 SKUs with at least 20 orders. When calculating the t-statistic, we use the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error.

Figure 4 plots the estimates of the warehouse and SKU regressions. The exhibit illustrates that the correlation between the *shipping time* and *customer lifetime* is significantly negative at the $p = 0.05$ level for 26 of 55 warehouses and 110 of 258 SKUs, and is significantly positive for only 1 warehouse and 1 SKU. Table 4 presents the estimates of the other regressions. It reports that the correlation between the *shipping time* and *customer lifetime* is significantly negative at the $p = 0.05$ level in 15 of 17 subsamples and significantly positive in no subsample.

6. Mechanisms

Newer customers could receive slower shipments for various reasons. For example, legacy customers could have more experience at identifying products that will be delivered fast. Or newer customers could be joining JD to buy more obscure items. Or JD could be prioritizing its legacy customers. Or legacy customers could be easier to serve because JD has more data about them. Or newer customers could be harder to reach (which in part explains their late adoption). We believe this last explanation is most likely. Indeed, Wong and Chao (2015) report that JD's newer customers are more heavily drawn from the countryside, as the company has largely tapped out its easy-to-reach customers in larger

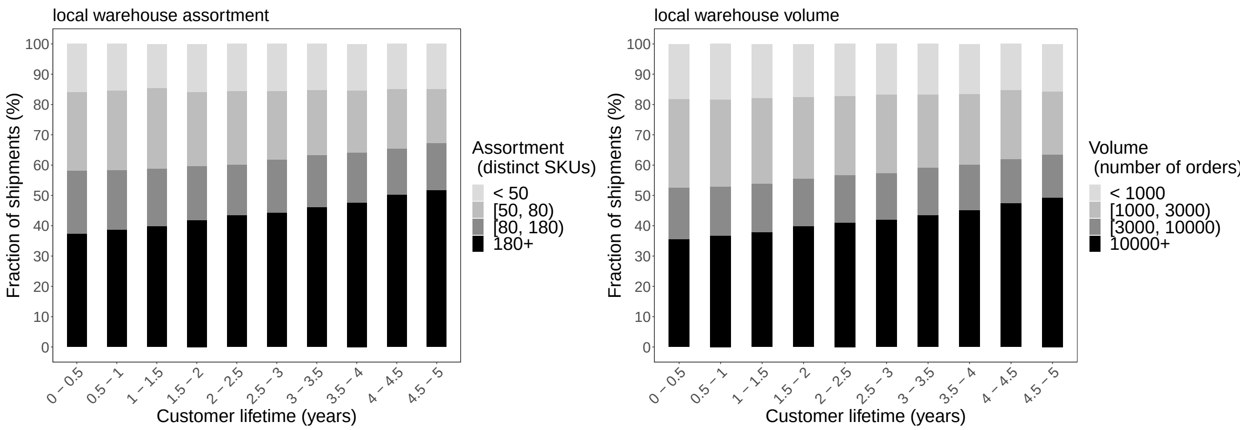

Figure 5 These plots show how the distributions of local warehouse assortment and local warehouse volume vary with customer lifetime across all shipments. For example, only 37.29% of orders made by customers who first purchased less than 6 months ago originated from the regions with local warehouses holding more than 180 distinct SKUs, in contrast with 50.2% of orders made by customers who first purchased 4-4.5 years ago.

cities. In addition, JD may have strategically built its warehouses near its established customer base rather than in the new centers of market growth.

6.1. Geographical expansion

We believe that newer customers receive slower shipments because they are harder to reach. To support this claim, we will show evidence that (1) newer customers tend to reside near smaller warehouses that are less capable of fulfilling local demand, and (2) even when newer customers are served by the local warehouse, it still takes longer to reach them.

Figure 5 illustrates that the local warehouses that are adjacent to newer customers tend to hold smaller assortments and fulfill fewer shipments. Indeed, OLS regressions of the *local warehouse assortment* and *local warehouse volume* on *customer lifetime* (while controlling for *active local warehouse*) suggest that increasing customer lifetime by one year increases the expected *local warehouse assortment* by about 5.15 and the expected *local warehouse volume* by about $1,022$. Both results are significant at the $p = 0.01$ level. The argument that newer customers tend to reside in areas with less-developed warehouses is also supported by Figure 6, which shows that such customers tend to reside in smaller cities.

Smaller size likely means that warehouses located near new customers are less capable of fulfilling local demand. To show that newer customers are indeed less likely to be served

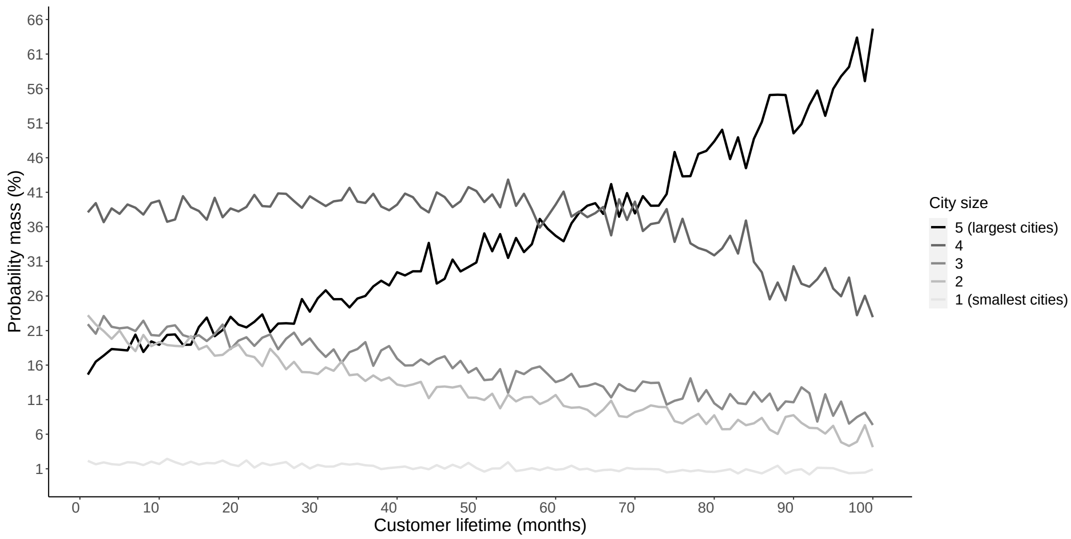

Figure 6 This figure plots the city size probability mass function by the number of months since the customer's first order.

by the nearest warehouse, we run an OLS regression of the *local shipping* on *customer lifetime* using the following specification:

$$local\ shipping\ = customer\ lifetime\ +\ controls, \tag{2}$$

where for controls, we use the *SKU id*, *active local warehouse*, *is_plus*, *quantity*, *order date*, and *order hour*.

We control for the *SKU id* because a difference in purchase behavior across customers with various lifetimes might also affect the probability of getting local shipping. The *active local warehouse* dummy variable allows us to disregard customers with local warehouses that did not ship orders.

As a robustness check, we also run the same logistic regression. We report the results in Table 5. We observe that the probability of getting a local shipment is positively correlated with the *customer lifetime*. This suggests that, after controlling for *active local warehouse*, newer customers are more likely to wait for shipments from faraway warehouses compared to legacy customers.

This fact, however, does not fully account for the shipping time discrepancy because local deliveries to new customers made from the closest warehouses also tend to take longer than local deliveries to legacy customers. To show this, we run the following OLS regression:

$$shipping\ time\ = customer\ lifetime\ +\ controls, \tag{3}$$

where for controls, we use the *local warehouse id*, *origin warehouse id*, *city size*, *SKU id*, *is_plus*, *quantity*, *order date*, and *order hour*.

We control for the *origin warehouse id* and *local warehouse id* to fix the shipment origin and destination and absorb any variation in the probability of local shipping. We also add the *city size* to absorb any fixed effects of city size. Finally, we add the *SKU id* to absorb differences in purchase behavior unrelated to the geographical location.

We report the results in Table 5. While the coefficient for the *customer lifetime* has predictably decreased after removing the variation in local shipping and location fixed effects (compared to Table 3), the *customer lifetime* and *shipping time* remain significantly negatively correlated. Since the *shipping time* is the time between when the package leaves the warehouse and when it reaches the customer, this result indicates that it takes longer to reach newer customers even when they are served by the same local warehouse as legacy customers.

Table 5 This table reports the customer lifetime estimates from the specifications (2) and (3) regressions. The numbers in parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error for the OLS regressions and the standard error for the logistic regression.

regression type	outcome variable	estimate
OLS	shipping time (hours)	-0.0155^{***} (0.00132)
OLS	local shipping (%)	0.121^{***} (0.00322)
logistic	local shipping	0.00888^{***} $(2.42 \cdot 10^{-4})$
Note:		*p<0.1; **p<0.05; ***p<0.01

While it is still possible that alternative mechanisms, such as a difference in purchase behavior or customer discrimination, could confound the effect of geographical expansion and contribute to slower deliveries to newer customers, we show in the next section that they are unlikely to play an instrumental role.

6.2. Alternative mechanisms

We investigate other possible mechanisms, such as a difference in purchase behavior between legacy and new customers or JD making shipping to legacy customers a higher priority. For example, legacy customers, given their experience with the platform, may strategically select items that are delivered faster. Or perhaps, JD rewards the loyalty of

Table 6 This table reports the customer lifetime estimates from the specification (4) regressions. The numbers in parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error for the OLS regression and the standard error for the logistic regression.

regression type	outcome variable	estimate
OLS	local shipping (%)	-0.0286^{***} (0.00334)
logistic	local shipping	-0.00228^{***} $(2.461 \cdot 10^{-4})$
Note:		*p<0.1; **p<0.05; ***p<0.01

legacy customers with faster shipping or is simply better at serving such customers because it has more information about their purchase behavior. These mechanisms, if they exist, could potentially contribute to slower deliveries to newer customers.

First, observe that the regression specification (3) already controls for the *SKU id* as well as for warehouses of origin and destination. Thus, we compare shipping times across customers purchasing the same SKUs on similar delivery routes. The persistent significant negative correlation between the *shipping time* and *customer lifetime* reported in Table 5 suggests that even if legacy customers are more strategic in their purchases, that fact is unlikely to play an instrumental role.

To strengthen our argument, we investigate whether customers with different lifetimes are more or less likely to get local shipping from the nearest warehouse. If legacy customers ordered faster-to-arrive items, we would expect them to order more items present in the warehouses closest to them than newer customers do. We run the following OLS regression of the *local shipping* on *customer lifetime*:

$$local\ shipping\ = customer\ lifetime\ +\ controls, \tag{4}$$

where for controls, we use the *local warehouse id* to absorb any variations between the local warehouses, and the *is_plus*, *is_gift*, *quantity*, *order date*, and *order hour*.

As a robustness check, we also run a logistic regression with the same specification. Table 6 shows that, when controlling for the *local warehouse id*, the probability of getting a local shipment is negatively correlated with the *customer lifetime*. This suggests that a given local warehouse is, in fact, less likely to ship to the nearest legacy customers than it is to ship to newer customers. In other words, legacy customers do not seem to strategically order items that are present in their local warehouses.

Recall that in the case of specification (2), without the *local warehouse id*, the sign of the coefficient for the *customer lifetime* was positive (Table 5); i.e., legacy customers were more likely to receive local shipping. This suggests that overall, legacy customers are more likely to reside near the larger warehouses and receive local shipping compared to newer customers who tend to reside around smaller warehouses. However, for a given local warehouse, legacy customers are not more likely to be served by that nearest warehouse than newer customers are.

Finally, it is reasonable to expect that if legacy customers were on average more strategic in ordering items that are faster to arrive, then legacy customers with a higher total value of past purchases would be even better at ordering such items. To test this idea, we regress the *shipping time* and *delivery time* on the interaction term between the *customer lifetime* and *purchase intensity*, *customer lifetime|purchase intensity*. The interaction term allows us to test whether legacy customers who purchased more in the past get even faster deliveries compared to legacy customers with the same lifetime who purchased less.

$$\begin{aligned} outcome\ variable\ &= customer\ lifetime\ +\ purchase\ intensity\ + \\ &\quad customer\ lifetime|purchase\ intensity\ +\ controls, \end{aligned} \tag{5}$$

where the *outcome variable* is either the *shipping time* or *delivery time*; and for controls, we use the *origin warehouse id*, *local warehouse id*, *city size*, *quantity*, *is_plus*, *is_gift*, *order date*, and *order hour*.

We add the *origin warehouse id*, *local warehouse id*, and *city size* to minimize the influence of location on delivery speed when comparing the levels of strategic behavior of customers.

Table 7 shows that the *shipping time* and *delivery time* in fact increase for legacy customers with the higher value of *purchase intensity*. This suggests that legacy customers with a higher total past purchase value are not better at ordering items that are faster to arrive compared to legacy customers with a lower total purchase value.

To conclude this section, we discuss whether JD may be making deliveries to legacy customers higher priorities, at the expense of newer customers. While our data do not allow us to test whether JD somehow expedites shipping to legacy customers, we can check whether JD dispatches the packages of legacy customers faster from the warehouses. Although not conclusive, it seems likely that if JD rushed to ship orders to legacy customers

Table 7 This table reports the interaction estimates from the specification (5) OLS regression relative to the customers with purchase intensity of 1. The numbers in parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error.

	shipping time (hours)	delivery time (hours)
customer lifetime|(purchase intensity = 2)	0.0096 (0.0376)	0.00481 (0.00845)
customer lifetime|(purchase intensity = 3)	0.0131* (0.00683)	0.0167** (0.00833)
customer lifetime|(purchase intensity = 4)	0.0153** (0.00647)	0.0217*** (0.00807)
customer lifetime|(purchase intensity = 0)	−0.124 (1.845)	−0.0299 (2.06)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 8 This table reports the customer lifetime estimate from the specification (6) OLS regression. The number in parentheses reports the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error.

outcome variable	estimate
order processing time (hours)	$-6.507 \cdot 10^{-4}$ (0.00108)
Note:	*p<0.1; **p<0.05; ***p<0.01

at the expense of newer customers, it would also serve the former faster at other stages of the delivery process.

Specifically, we use the outcome variable *order processing time*, and run the following OLS regression:

$$order\ processing\ time = customer\ lifetime\ +\ controls, \tag{6}$$

where for controls, we use the *origin warehouse id*, *local warehouse id*, *city size*, *SKU id*, *is_plus*, *quantity*, *order date*, and *order hour*.

We add the *SKU id*, *origin warehouse id*, *local warehouse id*, and *city size* variables to control for possible location-, and SKU-specific differences in the warehouse operations, allowing us to compare the *order processing time* for similar shipping routes and SKUs.

Our results, listed in Table 8, show that the *customer lifetime* is not significant, supporting visual observations in Figure 2 and suggesting that JD does not dispatch packages faster to legacy customers than to newer customers.

Table 9 This table reports the customer lifetime estimates from specification (1) regression for the *shipping time* as the outcome variable run across the 4 subsamples split by the promised delivery time. The numbers in parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error.

promised delivery time	estimate	observations
1 and 2 days	-0.0127^{***} (0.001)	193,767
1 day	-0.00426^{***} (0.00123)	115,774
2 days	-0.0151^{***} (0.00174)	77,993
3 days or longer	-0.0333^{***} (0.00836)	20,959
Note:	*p<0.1; **p<0.05; ***p<0.01	

Finally, Table 9 establishes that shipping speeds are slower for newer customers, even when we control for promised delivery times. Specifically, we find a strong negative correlation between *shipping time* and *customer lifetime* for all values of *promised delivery time*. Accordingly, our results are not fully explained by urban residents' greater demands for shorter promised delivery times.

7. Conclusion

The traditional advantages of economies of scale have long been considered a strength of large e-commerce platforms such as Amazon, Alibaba, and JD. However, our study in the context of JD highlights the potential negative logistical trade-offs that can arise from market growth. Specifically, we find that as JD expands into ever more remote areas, delivery times to newer customers become slower and less reliable. For example, we find that for each additional year a customer has been on the platform, the expected delivery time decreases by about 0.67 hours and the probability that packages do not arrive within 24 hours decreases by around 1.9%.

Our results suggest that the primary driver of this service quality heterogeneity is the gradual expansion of JD's distribution network into areas with geographically dispersed demand, where customers are harder to reach. This poses logistical difficulties that counteract the benefits of market growth and economies of scale, as it is not always possible for e-commerce firms to expand their supply chains as quickly as they expand their customer base. In particular, the warehouses closest to newer customers tend to hold smaller assortments and be less capable of fulfilling local demand.

Many e-commerce firms may face similar logistical trade-offs as they expand their operations and customer base, and it is important to carefully manage the synchronization between market growth and supply chain network expansion. In addition to improving delivery times, prioritizing the allocation of logistics resources to newer customers can also help e-commerce companies to lock-in these customers through higher-quality service. This is particularly important since newer customers are typically less loyal to the business compared to legacy customers, and they may be more likely to switch to a competitor if they experience poor service quality. Customer retention is critical for the long-term success and profitability of e-commerce platforms. As the online platforms market becomes increasingly competitive, firms must focus not only on acquiring new customers but also on retaining existing ones. Studies have shown that it can cost five times as much to acquire a new customer as it does to retain an existing one (Gallo, 2014), highlighting the importance of customer retention for reducing customer acquisition costs.

In conclusion, while economies of scale remain a critical strength of large e-commerce platforms, our study highlights the potential negative logistical trade-offs that can arise from market growth. By carefully managing the synchronization between market growth and supply chain network expansion, e-commerce companies can overcome these challenges and maintain sustainable growth and success in an increasingly competitive and dynamic market.

References

Anderson E, Simester D (2004) Long-Run Effects of Promotion Depth on New Versus Established Customers: Three Field Studies. *Marketing Science* 23(1):4–20.

Anupindi R, Chopra S, Deshmukh S, Van Mieghem J, Zemel E (2012) *Managing Business Process Flows: Principles of Operations Management* (New Jersey, USA: Prentice Hall).

Bray R (2020) Operational Transparency: Showing When Work Gets Done (Articles in advance). *Manufacturing & Service Operations Management* 1–15.

Cameron CA, Trivedi PK (2005) *Microeconometrics: Methods and Applications* (Cambridge, UK: Cambridge University Press).

Chen Y (1997) Paying Customers to Switch. *Journal of Economics and Management Strategy* 6(4):877–897.

Cui R, Li M, Li Q (2019) Value of High-Quality Logistics: Evidence from a Clash Between SF Express and Alibaba. *Management Science* 66(9):3879–3902.

Cui R, Lu M, Li Q (2021) Sooner or Later? Promising Delivery Speed in Online Retail. *Working paper* .

Daskin M (2013) *Network and Discrete Location: Models, Algorithms, and Applications* (New York, USA: John Wiley Sons, Inc.).

Deloitte (2015) The retail transformation. Cultivating choice, experience, and trust. *https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/consumer-business/DUP-955_Future-of-retail_vFINAL.PDF* Accessed: 2021-11-11.

Deloitte (2021) How brands can keep (or catch) digital marketplace momentum. *https://www.deloittedigital.com/us/en/blog-list/2021/how-brands-can-keep--or-catch--digital-marketplace-momentum.html* Accessed: 2021-11-10.

Eppen G (1979) Note—Effects of Centralization on Expected Costs in a Multi-Location Newsboy Problem. *Management Science* 25(5):498–501.

Farrell J, Klemperer P (2007) Coordination and lock-in: Competition with switching costs and network effects. Armstrong M, Porter R, eds., *Handbook of Industrial Organization*, volume 3, 1967–2072 (North-Holland).

Fisher ML, Gallino S, Xu JJ (2019) The Value of Rapid Delivery in Omnichannel Retailing. *Journal of Marketing Research* 56(5):732–748.

Gallo A (2014) The value of keeping the right customers. *Harvard business review* 29(10):304–309.

Gaur V, Kesavan S (2015) The effects of firm size and sales growth rate on inventory turnover performance in the u.s. retail sector. Agrawal N, Smith SA, eds., *Retail Supply Chain Management. International Series in Operations Research & Management Science*, volume 223, 25–52 (Boston, MA: Springer).

Giancotti M, Guglielmo A, Mauro M (2017) Efficiency and optimal size of hospitals: Results of a systematic search. *PLoS One* 12(3):1026–1041.

Holmes T (2011) The diffusion of Wal-Mart and economies of density. *Econometrica* 79(1):253–302.

Houde J, Newberry P, Seim H (2021a) Economies of Density in E-commerce: A Study of Amazon's Fulfillment Center Network. *NBER Working Paper* .

Houde J, Newberry P, Seim H (2021b) The Impact of Distance and Taxes on Amazon. *Working Paper* .

JD (2021) Financial and Operational Highlights. https://ir.jd.com/news-releases/news-release-details/jdcom-announces-fourth-quarter-and-full-year-2020-results Accessed: 2021-11-10.

Klemperer P (1995) Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade. *Review of Economic Studies* 62(4):515–539.

Lee H (2009) Online-Shopping Market in China - Adventurous Kingdom for Foreign SME. https://www.unifr.ch/intman/en/assets/public/intman/files/research/LEE_OnlineShopping_in_China.pdf Accessed: 2021-11-11.

Lim M, Mak H, Shen Z (2017) Agility and Proximity Considerations in Supply Chain Design. *Management Science* 63(4):1026–1041.

MacKinnon J, White H (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29(3):305–325.

Mandelbaum A, Reiman M (1998) On Pooling in Queueing Networks. *Management Science* 44(7):971–981.

Manne A (1961) Capacity Expansion and Probabilistic Growth. *Econometrika* 29(4):632–649.

Padilla A (1992) Mixed pricing in oligopoly with consumer switching costs. *International Journal of Industrial Organization* 10(3):393–411.

Pang G, Whitt W (2010) Service Interruptions in Large-Scale Service Systems. *Management Science* 55(9):1499–1512.

Rumyantsev S, Netessine S (2007) What Can Be Learned from Classical Inventory Models? A Cross-Industry Exploratory Investigation. *Manufacturing Service Operations Management* 9(4):409–429.

Shen M, Tang C, Wu D, Yuan R, Zhou W (2020) JD.com: Transaction Level Data for the 2020 MSOM Data Driven Research Challenge. *Manufacturing & Service Operations Management (Articles in Advance)* 1–9.

Shi Y, Yang Z, Yan H, Tian X (2017) Delivery Efficiency and Supplier Performance Evaluation in China's E-retailing Industry. *J Syst Sci Complex* 30:392–410.

Smith D, Whitt W (1981) Resource sharing for efficiency in traffic systems. *The Bell System Technical Journal* 60(1):39 – 55.

Song H, Tucker A, Murrell K (2015) The Diseconomies of Queue Pooling: An Empirical Investigation of Emergency Department Length of Stay. *Management Science* 61(12):3032–3053.

Watanabe S (2019) JD.com's profit crashes 80% in race with Alibaba over rural China. https://asia.nikkei.com/Business/Retail/JD.com-s-profit-crashes-80-in-race-with-Alibaba-over-rural-China Accessed: 2021-03-21.

Whitt W (1992) Understanding the Efficiency of Multi-Server Service Systems. *Management Science* 38(5):708–723.

Williamson O (1975) *Markets And Hierarchies: Analysis And Antitrust Implications* (New York, USA: The Free Press, Macmillan Publishing Co., Inc.).

Wong G, Chao L (2015) Alibaba, JD.com Target Rural China for E-Commerce Growth. https://www.wsj.com/articles/alibaba-jd-com-target-rural-china-for-e-commerce-growth-1440980597 Accessed: 2021-03-21.

WSJ (2021) Fuel Online Growth Through Digital Marketplaces. https://deloitte.wsj.com/articles/fuel-online-growth-through-digital-marketplaces-01625079609 Accessed: 2021-11-10.

Appendix. Robustness checks

A1. Customer lifetime as a categorical variable

We complement results in the paper by showing that slower service to newer customers propagates consistently throughout the customer lifetime. To do so, we define a new categorical variable, *lifetime group*, that aggregates customers into 6-month *customer lifetime* buckets. For example, customers have *lifetime group* = 1 if they signed up to the JD platform during the last 6 months, *lifetime group* = 2 for 6-12 months ago, *lifetime group* = 3 for 12-18 months ago, etc., until *lifetime group* = 12. We chose a 6-month interval to ensure a sufficient number of customers in each lifetime group.

We then run the following regressions on the *lifetime group*:

$$outcome\ variable\ = \sum_{i=2}^{12} lifetime\ group_i\ +\ controls, \tag{A.1}$$

where the *outcome variable* is either the *shipping time*, *delivery time*, or *promised delivery time* (for these three outcome variables, we run OLS regressions) or *multi-day delivery* (for which we run a logistic regression); for controls, we use the *is_plus*, *is_gift*, *quantity*, *order date*, and *order hour*.

Table A.1 and Figure A.1 (for the *shipping time*, *delivery time*, and *promised delivery time*) show that the *shipping time*, *delivery time*, *promised delivery time*, and the odds of waiting longer than a day for delivery all consistently decrease with each half a year of customer lifetime added, suggesting that newly acquired customers are progressively harder to reach.

A2. Middle mile shipping

We modify the regression with specification (3) to obtain a better approximation for the logistical complexity of shipping. First, we use a data sample with only local shipments to exclude possible unobservable confounding effects when shipments are made from distant warehouses, such as different wait times in intermediate warehouses along the way. Second, in addition to using the *shipping time*, we define a new outcome variable, *transit time*, which is the time between when the package leaves the warehouse and the time it arrives at the last-mile distribution facility; i.e., we exclude the time that the package spends in the last-mile delivery station as well as the last-mile delivery interval, which could also include various unobservable confounding effects.

Table A.1 This table reports the customer lifetime estimates from the specification (A.1) regressions. Each row shows the difference in the outcome variables between the given lifetime group and the newest customers, who first purchased less than 6 months ago. The numbers in parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error.

	shipping time (hours)	delivery time (hours)	promised delivery time (days)	multi-day delivery
customer lifetime: 0.5 - 1 years	−0.0892 (0.209)	−0.747*** (0.239)	0.00996 (0.00964)	−0.0453** (0.0197)
customer lifetime: 1 - 1.5 years	−0.497** (0.2)	−0.925*** (0.235)	−0.0235*** (0.00911)	−0.0611*** (0.0194)
customer lifetime: 1.5 - 2 years	−0.736*** (0.196)	−1.344*** (0.239)	−0.0388*** (0.00983)	−0.0939*** (0.0196)
customer lifetime: 2 - 2.5 years	−0.99*** (0.197)	−1.679*** (0.23)	−0.0662*** (0.00907)	−0.161*** (0.0195)
customer lifetime: 2.5 - 3 years	−1.643*** (0.199)	−2.211*** (0.235)	−0.0911*** (0.00934)	−0.217*** (0.0206)
customer lifetime: 3 - 3.5 years	−1.745*** (0.21)	−2.567*** (0.249)	−0.0989*** (0.0102)	−0.273*** (0.0217)
customer lifetime: 3.5 - 4 years	−2.189*** (0.222)	−3.206*** (0.26)	−0.139*** (0.0104)	−0.358*** (0.0233)
customer lifetime: 4 - 4.5 years	−2.584*** (0.24)	−3.613*** (0.265)	−0.172*** (0.0111)	−0.403*** (0.0238)
customer lifetime: 4.5 - 5 years	−2.499*** (0.235)	−3.706*** (0.282)	−0.185*** (0.0112)	−0.438*** (0.0254)
customer lifetime: 5 - 5.5 years	−3.244*** (0.227)	−4.189*** (0.281)	−0.233*** (0.011)	−0.506*** (0.0259)
customer lifetime: 5.5 - 6 years	−3.405*** (0.224)	−4.423*** (0.277)	−0.24*** (0.0111)	−0.541*** (0.0249)
Note:				*p<0.1; **p<0.05; ***p<0.01

We thus run the following OLS regressions:

$$outcome\ variable = customer\ lifetime + controls, \tag{A.2}$$

where the *outcome variable* is either the *shipping time* or *transit time*; for controls, we use the *origin warehouse id*, *city size*, *SKU id*, *is_plus*, *quantity*, *order date*, and *order hour* (we remove the *local warehouse id* because in the case of local shipments, it is equal to the *origin warehouse id*).

Table A.2 shows that the *customer lifetime* is significantly negatively correlated with the *shipping time* and *transit time*, supporting our argument that newer customers are, on average, harder to reach than legacy customers even when served by the same local warehouse.

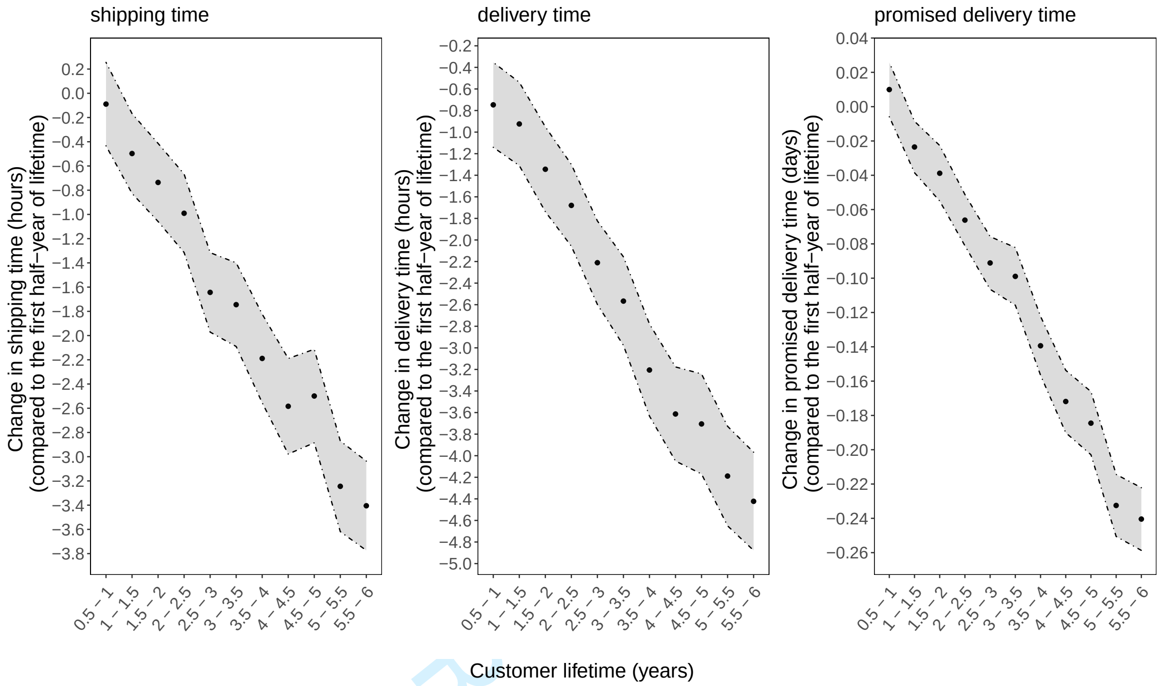

Figure A.1 This figure plots the regression coefficient estimates for a given customer lifetime group and the 90% confidence interval. From left to right, the outcome variables are the shipping time, delivery time, and promised delivery time. The y-axis indicates relative changes in all outcome variables compared to the newest customers who first purchased at most half a year ago. For example, the right-most point in the facet for the *shipping time* suggests that customers who first purchased less than 6 months ago, on average, get about 3.4 hours longer shipping time compared to customers who signed up between 5.5 and 6 years ago.

Table A.2 This table reports the customer lifetime estimates from the specification (A.2) regressions. The numbers in the parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error.

	shipping time (hours)	transit time (hours)
customer lifetime	-0.0146^{***} (0.00147)	-0.00777^{***} (0.000645)
Observations	140,285	140,285
Note:		*p<0.1; **p<0.05; ***p<0.01

A3. Data sample

Finally, we show that our results are robust to the data entries removed in Section 3.1. We restore all removed data except entries with missing promised delivery time and missing actual delivery information. We also remove 22 orders with non-JD SKUs that were shipped as multiple packages because we do not know what was included in each package.

Table A.3 This table reports the customer lifetime estimates from the specification (1) regressions with the restored data entries. The numbers in the parentheses report the maximum of the bootstrapped standard error and the heteroskedasticity robust standard error for the OLS regressions, and standard error for the logistic regressions.

regression type	outcome variable	without controls	with controls
OLS	shipping time (hours)	-0.108^{***} (0.00168)	-0.0939^{***} (0.0018)
OLS	delivery time (hours)	-0.117^{***} (0.00187)	-0.107^{***} (0.00202)
OLS	promised delivery time (days)	-0.00641^{***} ($7.529 \cdot 10^{-5}$)	-0.00578^{***} ($8.091 \cdot 10^{-5}$)
ordinal logistic regression	promised delivery time	-0.0131^{***} (0.00013)	-0.0123^{***} (0.000139)
OLS	multi-day delivery (days)	-0.218^{***} (0.00322)	-0.224^{***} (0.00337)
logistic	multi-day delivery	-0.00886^{***} (0.000134)	-0.00964^{***} (0.000148)
Observations		321,840	321,840
Note:		*p<0.1; **p<0.05; ***p<0.01	

This leaves $321,840$ rows and $291,539$ unique orders (since there could now be multiple SKUs per order). We repeat OLS regressions of the *shipping time*, *delivery time*, *promised delivery time*, and *multi-day delivery* on the *customer lifetime*; as well as the ordinal logistic regression of the *promised delivery time* on the *customer lifetime* and the logistic regression of the *multi-day delivery* on the *customer lifetime* using base specification (1). For the controls, we again use either nothing or the following five variables: *is_plus*, *is_gift*, *quantity*, *order date*, and *order hour*. Table A.3 shows that our previous results are robust to the restored data.